Volume 30, Issue 4

The Strange Case of Dr. “Unemployed” and Mr “Hidden” in Italy

Gaetano Lisi
University of Cassino

Abstract
While the literature is unanimous in considering unemployment and underground employment as strongly connected and interdependent phenomena, the link between existing causality is controversial. This empirical paper aims at clarifying this intricate relationship. Precisely, two key results emerge from the analysis. The first shows that unemployment (positively) influences the underground employment and vice versa. The second demonstrates that the impact of unemployment on underground employment is stronger than the reverse.
1. Introduction

The socio-economic dualism present in Italy is a universally recognised phenomenon (the so-called “Southern question”), which manifests itself most notably in relation to the underground economy. Italy, in fact, ranks as one of developed countries with the highest percentage of underground economy (Schneider et al., 2010) and is further characterised by the fact that two macro-areas with differing levels of shadow employment coexist within the same institutional structure (ISTAT, 2005, 2008a, 2008b). Furthermore, the highest levels of shadow employment present in the regions of Southern Italy are also respectively matched by the highest unemployment levels (Boeri and Garibaldi, 2002, 2006). Indeed, the literature is unanimous in considering underground employment and unemployment as strongly connected and interdependent phenomena. Boeri and Garibaldi go as far as to describe the two economical phenomena as « […] two sides of the same coin » (2006, p. 20).

However, the sign of the relation and above all the link between existing causality are controversial (Tanzi, 1999; Giles and Tedds, 2002). This is due to the fact that the shadow labour force is rather heterogeneous: a part of the shadow labour force is often incorrectly classified as unemployed (and is therefore included in the official labour force); another part is composed of retired people, minors, housewives and illegal immigrants that are not part of the labour force considered in the official statistics; and finally there are individuals that are part of both the official and shadow labour force, the so called “moonlighting” phenomenon (Tanzi, 1999).

Simple scatter diagrams are able to illustrate the tight positive relationship that exists between unemployment and shadow employment in Italy, both between regions (Figure 1) and over time (Figure 2). However, nothing can be deduced regarding the link between existing causality.

Basically, the tight relationship between unemployment and shadow employment is always hypothesised but is often not much analysed.

---

1 Consider the wide literature which has put the underground economy theory together with the benchmark macroeconomic model for the study of unemployment, i.e. the search and matching model (Bouev, 2002, 2005; Boeri and Garibaldi, 2002, 2006; Kolm and Larsen, 2003, 2010; Fugazza and Jacques, 2004; Albrecht et. al., 2009; Bosch and Esteban-Pretel, 2009).

2 Indeed, according to Bouev (2002, 2005), scaling down the unofficial sector can lead to a decrease in the level of unemployment; whereas according to Boeri and Garibaldi (2002, 2006), attempts to reduce shadow employment will result in higher open unemployment.

3 In short, the unemployed worker and the worker employed in the hidden sector can be the same person, as Dr Jekyll and Mr. Hide (Robert Louis Stevenson, 1886, The strange case of Dr Jekyll and Mr. Hide).
This regional panel analysis makes use of a Simultaneous Equations Model (SEM) and takes into account many of the factors that could influence both the unemployment and shadow employment rates, including the unobservable heterogeneity that is specific to each cross-section (regional) unit. Furthermore, considering the very tight relation between the two variables, suitable exogenous sources are introduced and the 3SLS (Three Stages Least Square) procedure is implemented.

The rest of the paper is organised as follows: section 2 presents the econometric model and the dataset; section 3 investigates the existing causality relationship between shadow employment and unemployment; while section 4 shows the results of the analysis; finally, section 5 concludes the work.

2. The econometric model

The panel used in this study is composed of 12 variables (cf. Table 1) obtained for the 20 Italian regions over 11 periods between 1995 and 2005, for a total of 220 observations. The two variables of interest – shadow employment and unemployment – are endogenous, i.e. they are simultaneously determined in equilibrium. Since both endogenous variables are fully observed, we thus have a typical Simultaneous Equations Model (Gujarati, 2003: Ch. 18-20):

\[
\begin{align*}
    h_{it} &= \beta_1 \cdot u_{it} + \alpha_1 \cdot Z_{it} + \mu_i + \zeta_{it} \\
    u_{it} &= \beta_2 \cdot h_{it} + \alpha_2 \cdot Z_{it} + \omega_i + \nu_{it}
\end{align*}
\]

hidden employment, \(h\), and unemployment, \(u\), are both explanatory and dependent variables. The fixed effects \(\mu_i\) and \(\omega_i\) account for all the unobservable variables and/or those not included in the analysis\(^6\) whereas, \(\zeta_{it}\) and \(\nu_{it}\) are the error terms\(^7\).

---

\(^4\) Although the relevance of economic-institutional factors (such as the excessive regulation and the tax burden) in accounting for the underground economy is widely accepted in the literature (see e.g. Schneider and Enste, 2000), the general view is that corruption, tax morality (or tax morale) and the poor quality of the institutions are equally significant factors (Tanzi, 1998; Johnson et al., 2000; Sarte, 2000; Fugazza and Jacques, 2004; Schneider, 2007; Torgler and Schneider, 2007).

\(^5\) The limited availability of data on underground employment determined this sample period.

\(^6\) According to Baltagi (2008), the fixed effects panel model is an adequate specification if the analysis is focalised on a specified group of \(N\) units and the inference is directed towards themselves. Furthermore, if the unobservable specific effects also represent omitted variables, it is highly likely that these effects are correlated with the other variables of the model, thus making the use of a fixed effects panel model indispensable (Judson and Owen, 1999). Finally, in panel applications characterised by a small number of temporal observations, it is standard procedure to use the more simple one-way individual specification (Baltagi, 2008).

\(^7\) In the Error Component Regression Models it is assumed that the regression disturbances are homoscedastic, with the same variance in time and between individuals. This hypothesis, although restrictive in some cases, will determine (in the case of heteroscedasticity) estimators that are always consistent but inefficient (Wooldridge, 2001; Baltagi, 2008). The same can be said about serial correlation.
The strong link existing between shadow employment and unemployment prevents a clear distinction between the determinants of each variable: as a consequence the control variables used in the analysis, $Z_u$, are the same in both the structural equations. The set of observables $Z_u$ comprises the main determinants of underground employment and unemployment. Following Daniele and Marani (2008), two variables are introduced as proxy of the size and regional economic structure: the synthetic index of infrastructure endowment, $\text{infr}$, and the rate of industrialisation, $\text{ind}$.

In order to determine the incidence of the organised crime, an index ($\text{ocr}$) defined by the sum of crimes typically charged to mafia type organisations (i.e. extortion and criminal organisation) for every 10,000 citizens was inserted into the model. As regards the corruption, an index ($\text{co}$) was created from the sum of the sentences executed for corruption, misappropriation (embezzlement), abuse of authority and acceptance of bribes for every 10,000 citizens.

Furthermore, amongst the control variables the following are included: the regional GDP per capita, $\text{gdp}$; the tax revenue collected by the Italian State in percentage of GDP (as a measure of tax burden), $\text{tax}$; an index relative to the regulation, $\text{reg}$; and the regional education level (the percentage of graduates and post graduates), $\text{istr}$.

A preliminary, and by no means exhaustive, analysis of the relations between the variables used in the model is given by the simple correlations (cf. Table 2).

3. Identification and Causality

It is believed that very few variables exist that are able to represent a source of exogeneity in the variation of the endogenous explicative and that are capable, therefore, of solving the problem of (incorrect) identification present in the model (1). Amongst these, two are particularly appropriate: i) the regional quota of illegal immigration ($\text{irr}$), as an instrument of the shadow rate; ii) the regional quota of socially useful workers ($\text{lsu}$), as an instrument of the unemployment rate.

---

8 Indeed, the important and negative role played by organised crime in the Italian Mezzogiorno must be kept in mind (Daniele and Marani, 2008; Marini and Turato, 2002).

9 Regarding the variable $\text{reg}$, the data is only available on a five-yearly basis and was therefore extended: for example, the data for 1995 was used for 1996, 1997, 1998 and 1999, until new data was available in 2000. Whereas, regarding the variable $\text{ind}$, the data is only available for the period 1997-2000 and for the year 2004; hence, the data for the period 1997-2000 was also used for 1995, 1996, 2001, 2002 and 2003, while the data for 2004 was also used for 2005.
In short, the fitted values of $h$ and $u$ to be used in the modified structural equations are obtained by the reduced form equations in which $lsu$ “replaces” $u$, and $irr$ “replaces” $h$, in such a way as to overcome the identification problem.

The reasons for the choice of $irr$ and $lsu$ are quite clear: illegal immigrants are not part of the official labour force and can only be used for shadow jobs, whereas socially useful workers are in actual fact unemployed individuals that work (albeit for fixed-term contracts and for specific projects) and have no interest/advantage (or the time) to work in the hidden sector.

Following Bianchi et al. (2008), both regularisation applications and permits of stay are used to identify the hidden component of immigration, taking advantage of the strong, positive and stable correlation in time and space between the legal and illegal immigration quotas (Bianchi et al., 2008). Indeed, Bianchi et al. (2008) show that the discrepancy between the actual illegal immigration quota and that estimated through the permits of stay is almost negligible.

As regards the socially useful workers, in order to aid the mapping of the phenomenon at a regional level, only provincial and regional projects were referred to.\textsuperscript{10}

Both the “found” ($irr$, $lsu$) and “created” ($\hat{h}$, $\hat{u}$) instruments adequately explain the endogenous variable to be instrumented ($h$ and $u$, respectively). The univariate regressions, implemented for this purpose, confirm that the chosen instruments are not weak: the first-stage $F$-statistics of the regressions, used to investigate whether a instrument is weak or not, is in fact greater than the lower limit indicated in the literature (Shea, 1997; Godfrey, 1999; Stock and Yogo, 2002; Andrews and Stock, 2005).\textsuperscript{11} Precisely, the first-stage $F$-statistics of the univariate regressions are equal to 14.78 ($h$ on $irr$), 15.14 ($u$ on $lsu$), 13.52 ($h$ on $\hat{h}$), and 14.24 ($u$ on $\hat{u}$).

Finally, considering that the set of control variables for $h$ and $u$ is basically the same, the Three Stages Least Square (3SLS) procedure is applied.\textsuperscript{12} More precisely, the reduced

\textsuperscript{10} The difficulty in obtaining data derives from the fact that the projects for these work typologies can also be promoted and carried out by municipalities, in other words by the smaller public administrations.

\textsuperscript{11} In order to investigate if an instrument is weak or not, Stock and Yogo (2002) developed tests based on the $F$-statistics, under the null hypothesis that the coefficient associated with the instrument is null in the univariate regression. More precisely, the $F$-test rejects the null hypothesis of a weak instrument, at the confidence level of 5%, if $F > 10.3$ (Andrews and Stock, 2005).

\textsuperscript{12} With respect to the Two stages approach, this procedure has the advantage of being more efficient due to the fact that the correlation between errors of the two structural equations is taken into account.
form equations are estimated using Within estimation, whereas the modified structural equations are estimated via SUR (Seemingly Unrelated Regression).

4. Results of the analysis
The estimation obtained by the model (1) – see Table 3 – shows a very interesting albeit not surprising result that clarifies the investigated causality relationship: unemployment (positively) influences the underground employment and vice versa. The coefficients are, in fact, both significant and positive. Therefore,

**Remark 1.** The causal relationship that links the unemployment rate with the underground employment rate appears to be bidirectional. Furthermore, this relationship appears to be asymmetric, since the effect of $u$ on $h$ is “stronger” than that of $h$ on $u$.

More precisely, an increase in unemployment of 1% is associated with an increase of 0.69% in underground employment. Vice versa, an increase of 1% in shadow employment is associated with an increase of 0.18% in unemployment. A possible economic explanation of this result is the following: an increase in the unemployment rate makes a higher manpower available to the underground sector; whereas, an increase in the underground employment may imply a reduction of official jobs, thus increasing the “official” unemployment rate recorded by the government.

In addition, several interesting remarks can be made:

- The variables corruption and organised crime are significant (and positive) only with respect to the dependent variable shadow employment. This result confirms the lack of univocal conclusions in the empirical literature that investigates the relationship between unemployment and organised crime (Marselli and Vannini, 2000), but also emphasises the strong relationship between these factors and the spread of the underground sector. These results can be easily extended to the European context, since corruption and organised crime are particularly widespread in the Eastern European countries where the underground economy is higher than in the rest of Europe (Van Dijk, 2006; Johnson et al., 2000);
• The GDP is significant only in reference to the dependent variable \( u \) (obviously negative in sign). This confirms the ambiguity in the literature regarding the relationship between GDP and the underground economy (Eilat and Zinnes, 2000);

• Education is statistically significant only in the reduction of underground employment. This result has recently been confirmed through empirical studies in Italy (Cappariello and Zizza, 2009).

• The variables \( tax \) and \( reg \) are statistically significant (with positive sign) only in reference to unemployment; whereas the significance of \( ind \) (with negative sign), with respect to both \( u \) and \( h \), means that the more industrialized is the region, the lower is unemployment and underground employment.

Finally, it can state that:

**Remark 2.** The strong difference between the North and the South of Italy, in terms of shadow economy, crucially depends on the different level of corruption and organized crime. Indeed, the southern regions of Italy constitute a typical case in which the socio-economic context of organized crime (Peri, 2004; Daniele and Marani, 2008, see also Figure 3), and of “amoral familism” (Banfield, 1958) has heavily burdened the economy.

5. Conclusions

The tight relationship between unemployment and shadow employment is always hypothesised but is often not much analysed, due to objective and non trivial difficulties (that have been briefly discussed in the introduction). For this reason the bidirectional causality result obtained, although not surprising, contributes in clarifying this very important topic.

A possible and interesting extension of this analysis could be based on the hypothesis that causality, although bidirectional, is not simultaneous. Essentially, unemployment and shadow employment influence each other respectively, but not simultaneously. This could be investigated by implementing a panel VAR methodology. However, the availability of a panel dataset with additional temporal observations is essential.
References


FIGURES AND TABLES

Figure 1. Underground employment and unemployment in the regions of Italy
(Source: Boeri and Garibaldi, 2006)

Figure 2. Underground employment and unemployment in the time in Italy
(Source: Boeri and Garibaldi, 2006)
Figure 3. Underground employment and organized crime in the regions of Italy
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u )</td>
<td>Regional unemployment rate (in %)</td>
<td>ISTAT</td>
</tr>
<tr>
<td>( h )</td>
<td>Regional underground employment rate (ratio between the regional underground employment and the regional total employment)</td>
<td>ISTAT</td>
</tr>
<tr>
<td>( irr )</td>
<td>Number of illegal immigrants for every 10,000 citizens (regional rate)</td>
<td>Italian Ministry of Interior</td>
</tr>
<tr>
<td>( lsu )</td>
<td>Number of socially useful workers for every 10,000 citizens (regional rate)</td>
<td>INPS (National Social Security Institute), Regions, Provinces</td>
</tr>
<tr>
<td>( gdp )</td>
<td>Regional GDP per capita</td>
<td>ISTAT</td>
</tr>
<tr>
<td>( infr )</td>
<td>Synthetic index of regional infrastructural endowment (Italy index = 100)</td>
<td>Unioncamere and “Tagliacarne” Institute</td>
</tr>
<tr>
<td>( ind )</td>
<td>Rate of industrialisation (ratio between the regional total employment in the manufacturing sector and the regional total employment)</td>
<td>Our elaboration on ISTAT data</td>
</tr>
<tr>
<td>( istr )</td>
<td>Regional education level (the percentage of graduates and post graduates on the regional resident population above 15 years of age)</td>
<td>ISTAT</td>
</tr>
<tr>
<td>( co )</td>
<td>Sum of the sentences executed for corruption, misappropriation (embezzlement), abuse of authority and acceptance of bribes for every 10,000 citizens (regional rate)</td>
<td>Our elaboration on Judicial Register data</td>
</tr>
<tr>
<td>( ocr )</td>
<td>Sum of crimes typically charged to mafia type organisations (i.e. extortion and criminal organisation) for every 10,000 (regional rate)</td>
<td>Our elaboration on Geographical Information System on Justice</td>
</tr>
<tr>
<td>( tax )</td>
<td>Tax revenue collected by the Italian State in percentage of GDP</td>
<td>OECD</td>
</tr>
<tr>
<td>( reg )</td>
<td>Regulation index (Italy)</td>
<td>OECD</td>
</tr>
</tbody>
</table>
Table 2. Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>h</th>
<th>gdp</th>
<th>infr</th>
<th>ind</th>
<th>u</th>
<th>co</th>
<th>ocr</th>
<th>tax</th>
<th>reg</th>
<th>istr</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gdp</td>
<td>-0.8470</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>infr</td>
<td>-0.3251</td>
<td>0.3401</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ind</td>
<td>-0.6418</td>
<td>0.4447</td>
<td>0.0975</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>0.8737</td>
<td>-0.9112</td>
<td>-0.3027</td>
<td>-0.5797</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>co</td>
<td>0.1700</td>
<td>-0.3206</td>
<td>0.1107</td>
<td>-0.0675</td>
<td>0.2020</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ocr</td>
<td>0.7666</td>
<td>-0.7989</td>
<td>-0.2101</td>
<td>-0.3495</td>
<td>0.7917</td>
<td>0.2028</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tax</td>
<td>0.1079</td>
<td>-0.0622</td>
<td>0.0000</td>
<td>0.0320</td>
<td>-0.010</td>
<td>0.2864</td>
<td>-0.0226</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reg</td>
<td>0.1301</td>
<td>-0.3235</td>
<td>0.0000</td>
<td>0.0254</td>
<td>0.0756</td>
<td>0.4250</td>
<td>-0.0158</td>
<td>0.1706</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>istr</td>
<td>-0.5235</td>
<td>0.6012</td>
<td>0.5139</td>
<td>0.1872</td>
<td>0.5550</td>
<td>0.1321</td>
<td>-0.3430</td>
<td>-0.3225</td>
<td>-0.2861</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

*Note: the numbers reported are the value (with the relative sign) and significance (p-value) of the correlation.*
Table 3. Results from 3SLS estimation

<table>
<thead>
<tr>
<th></th>
<th>$h$</th>
<th>$u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td>0.69 (8.65) [0.000]</td>
<td>-</td>
</tr>
<tr>
<td>$h$</td>
<td>-</td>
<td>0.18 (5.72) [0.000]</td>
</tr>
<tr>
<td>gdp</td>
<td>0.08 (1.39) [0.167]</td>
<td>-0.18 (-2.25) [0.025]</td>
</tr>
<tr>
<td>infr</td>
<td>-0.45 (-1.19) [0.235]</td>
<td>-0.61 (-1.09) [0.274]</td>
</tr>
<tr>
<td>ind</td>
<td>-0.20 (-7.08) [0.000]</td>
<td>-0.17 (-4.32) [0.000]</td>
</tr>
<tr>
<td>istr</td>
<td>-0.44 (-2.70) [0.007]</td>
<td>-0.51 (-1.37) [0.173]</td>
</tr>
<tr>
<td>co</td>
<td>0.25 (5.84) [0.000]</td>
<td>0.09 (0.72) [0.469]</td>
</tr>
<tr>
<td>ocr</td>
<td>0.39 (5.70) [0.000]</td>
<td>0.13 (0.71) [0.479]</td>
</tr>
<tr>
<td>reg</td>
<td>0.30 (1.62) [0.106]</td>
<td>0.34 (1.80) [0.072]*</td>
</tr>
<tr>
<td>tax</td>
<td>0.15 (1.54) [0.125]</td>
<td><strong>0.29</strong> (2.57) [0.010]</td>
</tr>
</tbody>
</table>

**Obs.** 220 220  
$R^2$ 0.3251 0.2150  
$\chi^2$ 572.71 [0.0000] 478.65 [0.0000]

**Breusch-Pagan test of independence:** $\chi^2(1) = 7.865$, $Pr = 0.0050$

(null hypothesis: no correlation between the error terms of the two modified structural equations)

**Notes:** The reduced form equations are estimated equation by equation, whereas the modified structural equations are estimated via SUR (Seemingly Unrelated Regression); indeed, 3SLS = 2SLS + SUR. All variables are expressed in log and are defined in Table 1. The numbers in round brackets are the z-ratios, whereas the numbers in square brackets are the p-value. The numbers in bold denote significance at 5% level, whereas * denotes significance only at 10% level.